

Обучение машинного перевода без параллельных текстов*

Рябов Ф. А. Бахтеев¹ О. Ю. Стрижов² В. В.

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

В работе рассматривается задача построения системы машинного перевода. В стандартном подходе к такой задаче используется набор параллельных предложений, состоящих из пар предложений, являющихся переводом друг друга. Однако в таком случае система нуждается в огромном объеме тестовой выборки, исчисляемой миллионами. В данной задаче рассматривается метод построения модели машинного перевода без использования параллельного корпуса. Основой метода является Seq2Seq, где предложениям каждого языка ставится в соответствие вектор из общего векторного пространства. Вычислительный эксперимент проводится на паре "русский-украинский".

Ключевые слова: *машинный перевод, нейросеть, Seq2Seq.*

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов, Eric Gaussier, координатор Малиновский Г. С. Консультант: Бахтеев О. Ю.